

OIPE

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/787,737

DATE: 07/19/2001

TIME: 12:21:32

Input Set : A:\001560-393.ST25.txt

Output Set: N:\CRF3\07192001\I787737.raw

ENTERED

4 <110> APPLICANT: Kakimoto, Tatsuo
 6 <120> TITLE OF INVENTION: Homoeobox Gene Encoding a Protein Involved in Differentiation
 8 <130> FILE REFERENCE: 001560-393
 10 <140> CURRENT APPLICATION NUMBER: US 09/787,737
 C--> 11 <141> CURRENT FILING DATE: 2001-06-26
 13 <150> PRIOR APPLICATION NUMBER: PCT/JP00/04904
 14 <151> PRIOR FILING DATE: 2000-07-21
 16 <150> PRIOR APPLICATION NUMBER: JP 11-207995
 17 <151> PRIOR FILING DATE: 1999-07-22
 19 <160> NUMBER OF SEQ ID NOS: 8
 21 <170> SOFTWARE: PatentIn version 3.0
 24 <210> SEQ ID NO: 1
 25 <211> LENGTH: 1214
 26 <212> TYPE: DNA
 27 <213> ORGANISM: Arabidopsis thaliana
 29 <220> FEATURE:
 30 <221> NAME/KEY: CDS
 31 <222> LOCATION: (36)..(1010)
 32 <223> OTHER INFORMATION: Nucleotide sequence coding for a protein involved in
 33 differentiation
 36 <400> SEQUENCE: 1
 37 ctttagctct cgattatcat cattacacca tcattc atg tcc tcc tca aac aaa 53
 38 Met Ser Ser Ser Asn Lys
 39 1 5
 41 aat tgg cca agc atg ttc aaa tcc aaa cct tgc aac aat aat cat cat 101
 42 Asn Trp Pro Ser Met Phe Lys Ser Lys Pro Cys Asn Asn Asn His His
 43 10 15 20
 45 cat caa cat gaa atc gat act cca tct tac atg cac tac tct aat tgc 149
 46 His Gln His Glu Ile Asp Thr Pro Ser Tyr Met His Tyr Ser Asn Cys
 47 25 30 35
 49 aac cta tca tct tcc ttt tcc tca gat cgg ata cca gat cct aaa ccg 197
 50 Asn Leu Ser Ser Ser Phe Ser Ser Asp Arg Ile Pro Asp Pro Lys Pro
 51 40 45 50
 53 aga tgg aat cct aaa ccg gag cag att agg ata ctc gaa tca atc ttc 245
 54 Arg Trp Asn Pro Lys Pro Glu Gln Ile Arg Ile Leu Glu Ser Ile Phe
 55 55 60 65 70
 57 aat tcc ggt act att aac cca cct aga gag gag att caa aga atc ccg 293
 58 Asn Ser Gly Thr Ile Asn Pro Pro Arg Glu Glu Ile Gln Arg Ile Arg
 59 75 80 85
 61 atc cgg ctt caa gaa tat ggt caa atc ggt gac gca aac gtg ttt tac 341
 62 Ile Arg Leu Gln Glu Tyr Gly Gln Ile Gly Asp Ala Asn Val Phe Tyr
 63 90 95 100
 65 tgg ttt caa aac ccg aaa tct cga gca aaa cac aag ctt cgt gtt cat 389
 66 Trp Phe Gln Asn Arg Lys Ser Arg Ala Lys His Lys Leu Arg Val His
 67 105 110 115
 69 cac aaa agc cct aaa atg tca aag aag gac aag acg gtt att cct agt 437
 70 His Lys Ser Pro Lys Met Ser Lys Lys Asp Lys Thr Val Ile Pro Ser

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/787,737

DATE: 07/19/2001

TIME: 12:21:32

Input Set : A:\001560-393.ST25.txt

Output Set: N:\CRF3\07192001\I787737.raw

```

71      120      125      130
73 act gac gct gat cat tgt ttt ggt ttt gtt aac caa gaa acc gga tta      485
74 Thr Asp Ala Asp His Cys Phe Gly Phe Val Asn Gln Glu Thr Gly Leu
75 135      140      145      150
77 tat ccg gtt caa aac aat gag ttg gtg gta acc gaa ccg gcc ggt ttt      533
78 Tyr Pro Val Gln Asn Asn Glu Leu Val Val Thr Glu Pro Ala Gly Phe
79      155      160      165
81 cta ttt ccg gtt cat aat gat ccg agc gct gct caa tca gcg ttt ggt      581
82 Leu Phe Pro Val His Asn Asp Pro Ser Ala Ala Gln Ser Ala Phe Gly
83      170      175      180
85 ttt ggc gat ttt gtt gta ccg gtg gta acg gaa gaa ggg atg gca ttc      629
86 Phe Gly Asp Phe Val Val Pro Val Val Thr Glu Glu Gly Met Ala Phe
87      185      190      195
89 tct acc gtt aat aac ggc gtt aat ttg gag act aac gaa aat ttt gat      677
90 Ser Thr Val Asn Asn Gly Val Asn Leu Glu Thr Asn Glu Asn Phe Asp
91      200      205      210
93 aaa att ccg gcg atc aat tta tac ggc gga gat gga aat ggc ggt gga      725
94 Lys Ile Pro Ala Ile Asn Leu Tyr Gly Gly Asp Gly Asn Gly Gly Gly
95 215      220      225      230
97 aat tgt ttt cct cct ttg act gtt cca tta acc atc aat caa tct caa      773
98 Asn Cys Phe Pro Pro Leu Thr Val Pro Leu Thr Ile Asn Gln Ser Gln
99      235      240      245
101 gaa aaa cga gat gta gga tta tcc ggt ggt gaa gac gtc gga gat aat      821
102 Glu Lys Arg Asp Val Gly Leu Ser Gly Gly Glu Asp Val Gly Asp Asn
103      250      255      260
105 gtt tat ccg gtg aga atg acg gtg ttt att aac gag atg cct atc gaa      869
106 Val Tyr Pro Val Arg Met Thr Val Phe Ile Asn Glu Met Pro Ile Glu
107      265      270      275
109 gta gtg tct gga tta ttc aac gtt aag gca gct ttc gga aac gat gcc      917
110 Val Val Ser Gly Leu Phe Asn Val Lys Ala Ala Phe Gly Asn Asp Ala
111      280      285      290
113 gtt ttg atc aac tcg ttt ggc cag cct att ctt aca gat gaa ttt ggt      965
114 Val Leu Ile Asn Ser Phe Gly Gln Pro Ile Leu Thr Asp Glu Phe Gly
115 295      300      305      310
117 gtt act tat caa cct ctc caa aat ggc gca atc tat tat ctt att      1010
118 Val Thr Tyr Gln Pro Leu Gln Asn Gly Ala Ile Tyr Tyr Leu Ile
119      315      320      325
121 tagaagatat tgaaaagcaa atgttatggt gctatggata aatattaata taataataaa      1070
123 agattttctgc gatttattta gttattaatt agataagaat ttcattttctt atcttttaaa      1130
125 tttatgaaca atttacagga catttacatt ttcgagactt tgaaaaataa agaatagaat      1190
127 taagttaaaaa aaaaaaaaaa aaaa      1214
131 <210> SEQ ID NO: 2
132 <211> LENGTH: 325
133 <212> TYPE: PRT
134 <213> ORGANISM: Arabidopsis thaliana
136 <220> FEATURE:
137 <223> OTHER INFORMATION: Amino acid sequence coding for a protein involved in
138 differentiation
140 <400> SEQUENCE: 2

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/787,737

DATE: 07/19/2001

TIME: 12:21:32

Input Set : A:\001560-393.ST25.txt

Output Set: N:\CRF3\07192001\I787737.raw

```

142 Met Ser Ser Ser Asn Lys Asn Trp Pro Ser Met Phe Lys Ser Lys Pro
143 1 5 10 15
146 Cys Asn Asn Asn His His His Gln His Glu Ile Asp Thr Pro Ser Tyr
147 20 25 30
150 Met His Tyr Ser Asn Cys Asn Leu Ser Ser Ser Phe Ser Ser Asp Arg
151 35 40 45
154 Ile Pro Asp Pro Lys Pro Arg Trp Asn Pro Lys Pro Glu Gln Ile Arg
155 50 55 60
158 Ile Leu Glu Ser Ile Phe Asn Ser Gly Thr Ile Asn Pro Pro Arg Glu
159 65 70 75 80
162 Glu Ile Gln Arg Ile Arg Ile Arg Leu Gln Glu Tyr Gly Gln Ile Gly
163 85 90 95
166 Asp Ala Asn Val Phe Tyr Trp Phe Gln Asn Arg Lys Ser Arg Ala Lys
167 100 105 110
170 His Lys Leu Arg Val His His Lys Ser Pro Lys Met Ser Lys Lys Asp
171 115 120 125
174 Lys Thr Val Ile Pro Ser Thr Asp Ala Asp His Cys Phe Gly Phe Val
175 130 135 140
178 Asn Gln Glu Thr Gly Leu Tyr Pro Val Gln Asn Asn Glu Leu Val Val
179 145 150 155 160
182 Thr Glu Pro Ala Gly Phe Leu Phe Pro Val His Asn Asp Pro Ser Ala
183 165 170 175
186 Ala Gln Ser Ala Phe Gly Phe Gly Asp Phe Val Val Pro Val Val Thr
187 180 185 190
190 Glu Glu Gly Met Ala Phe Ser Thr Val Asn Asn Gly Val Asn Leu Glu
191 195 200 205
194 Thr Asn Glu Asn Phe Asp Lys Ile Pro Ala Ile Asn Leu Tyr Gly Gly
195 210 215 220
198 Asp Gly Asn Gly Gly Gly Asn Cys Phe Pro Pro Leu Thr Val Pro Leu
199 225 230 235 240
202 Thr Ile Asn Gln Ser Gln Glu Lys Arg Asp Val Gly Leu Ser Gly Gly
203 245 250 255
206 Glu Asp Val Gly Asp Asn Val Tyr Pro Val Arg Met Thr Val Phe Ile
207 260 265 270
210 Asn Glu Met Pro Ile Glu Val Val Ser Gly Leu Phe Asn Val Lys Ala
211 275 280 285
214 Ala Phe Gly Asn Asp Ala Val Leu Ile Asn Ser Phe Gly Gln Pro Ile
215 290 295 300
218 Leu Thr Asp Glu Phe Gly Val Thr Tyr Gln Pro Leu Gln Asn Gly Ala
219 305 310 315 320
222 Ile Tyr Tyr Leu Ile
223 325
227 <210> SEQ ID NO: 3
228 <211> LENGTH: 1518
229 <212> TYPE: DNA
230 <213> ORGANISM: Arabidopsis thaliana
232 <220> FEATURE:
233 <221> NAME/KEY: CDS
234 <222> LOCATION: (152)..(1285)

```

RAW SEQUENCE LISTING

DATE: 07/19/2001

PATENT APPLICATION: US/09/787,737

TIME: 12:21:32

Input Set : A:\001560-393.ST25.txt

Output Set: N:\CRF3\07192001\I787737.raw

235 <223> OTHER INFORMATION: Nucleotide sequence coding for a protein involved in
 236 differentiation
 238 <400> SEQUENCE: 3
 239 tttttattta tctttccttt agccattctg ttccctgtct cttcctctc tctttttgac 60
 241 acatcacatc atcatcacat catcattcaa catcaatcat catcatatgc atacacatac 120
 243 atctgtgttc tgcgcatcga gtttaattagt t atg gct tct tgc aat aga cac 172
 244 Met Ala Ser Ser Asn Arg His
 245 1 5
 247 tgg cca agc atg ttc aag tcc aaa cct cat ccc cat caa tgg caa cat 220
 248 Trp Pro Ser Met Phe Lys Ser Lys Pro His Pro His Gln Trp Gln His
 249 10 15 20
 251 gac atc aac tct cct ctc ttg cct tct gct tct cac cga tct tct cct 268
 252 Asp Ile Asn Ser Pro Leu Leu Pro Ser Ala Ser His Arg Ser Ser Pro
 253 25 30 35
 255 ttc tct tca gga tgt gaa gtg gag agg agt cca gag cca aaa cca aga 316
 256 Phe Ser Ser Gly Cys Glu Val Glu Arg Ser Pro Glu Pro Lys Pro Arg
 257 40 45 50 55
 259 tgg aat cca aag cca gag cag att cgg ata ctt gaa gca atc ttt aac 364
 260 Trp Asn Pro Lys Pro Glu Gln Ile Arg Ile Leu Glu Ala Ile Phe Asn
 261 60 65 70
 263 tcc ggg atg gtg aat cct cca aga gag gag atc agg agg att agg gct 412
 264 Ser Gly Met Val Asn Pro Pro Arg Glu Glu Ile Arg Arg Ile Arg Ala
 265 75 80 85
 267 cag ctt caa gaa tac ggc caa gtc ggt gat gct aac gtc ttc tac tgg 460
 268 Gln Leu Gln Glu Tyr Gly Gln Val Gly Asp Ala Asn Val Phe Tyr Trp
 269 90 95 100
 271 ttc caa aac cgt aag tcc cgt agt aaa cac aaa ctc cgc ctc ctc cac 508
 272 Phe Gln Asn Arg Lys Ser Arg Ser Lys His Lys Leu Arg Leu Leu His
 273 105 110 115
 275 aac cac tcc aaa cac tct ctc cct caa acg caa ccg cag ccg cag ccg 556
 276 Asn His Ser Lys His Ser Leu Pro Gln Thr Gln Pro Gln Pro Gln Pro
 277 120 125 130 135
 279 caa cct tgc gct tcc tct tcc tct tcc tcc tct tcc tcc tcc aaa 604
 280 Gln Pro Ser Ala Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Lys
 281 140 145 150
 283 tcc acc aaa ccc cga aaa agc aag aac aag aac aac act aat ctc tct 652
 284 Ser Thr Lys Pro Arg Lys Ser Lys Asn Lys Asn Asn Thr Asn Leu Ser
 285 155 160 165
 287 ttg ggt ggt agt caa atg atg ggg atg ttt cca ccg gaa ccg gcg ttt 700
 288 Leu Gly Gly Ser Gln Met Met Gly Met Phe Pro Pro Glu Pro Ala Phe
 289 170 175 180
 291 ctc ttc ccg gtc tcc act gtc gga ggg ttt gaa ggt atc acc gtc tca 748
 292 Leu Phe Pro Val Ser Thr Val Gly Gly Phe Glu Gly Ile Thr Val Ser
 293 185 190 195
 295 tcc caa tta ggg ttt ctc tcc ggt gat atg att gag caa caa aaa ccg 796
 296 Ser Gln Leu Gly Phe Leu Ser Gly Asp Met Ile Glu Gln Gln Lys Pro
 297 200 205 210 215
 299 gct cca acg tgt acc gga ctc ctg ctg agt gag atc atg aac ggt agt 844
 300 Ala Pro Thr Cys Thr Gly Leu Leu Leu Ser Glu Ile Met Asn Gly Ser

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/787,737

DATE: 07/19/2001

TIME: 12:21:32

Input Set : A:\001560-393.ST25.txt

Output Set: N:\CRF3\07192001\I787737.raw

```

301          220          225          230
303 gtg agt tat gga act cat cat caa caa cac ttg agt gag aaa gaa gtt      892
304 Val Ser Tyr Gly Thr His His Gln Gln His Leu Ser Glu Lys Glu Val
305          235          240          245
307 gaa gaa atg agg atg aag atg ttg caa cag cca cag act cag att tgt      940
308 Glu Glu Met Arg Met Lys Met Leu Gln Gln Pro Gln Thr Gln Ile Cys
309          250          255          260
311 tac gct acc act aat cat caa ata gct tct tac aac aac aac aac aac      988
312 Tyr Ala Thr Thr Asn His Gln Ile Ala Ser Tyr Asn Asn Asn Asn Asn
313          265          270          275
315 aac aat aac atc atg ctt cat att cct ccc act act tct act gcc acc      1036
316 Asn Asn Asn Ile Met Leu His Ile Pro Pro Thr Thr Ser Thr Ala Thr
317 280          285          290          295
319 act att act act tcg cat tct ctc gct act gtc cca tca act tcg gac      1084
320 Thr Ile Thr Thr Ser His Ser Leu Ala Thr Val Pro Ser Thr Ser Asp
321          300          305          310
323 cag ctt caa gtt caa gcg gac gca cga ata aga gtt ttc atc aat gaa      1132
324 Gln Leu Gln Val Gln Ala Asp Ala Arg Ile Arg Val Phe Ile Asn Glu
325          315          320          325
327 atg gag ctt gaa gtg agc tca gga ccg ttc aat gtg agg gat gca ttt      1180
328 Met Glu Leu Glu Val Ser Ser Gly Pro Phe Asn Val Arg Asp Ala Phe
329          330          335          340
331 ggg gaa gag gtt gtt ctg att aat tcc gcg ggt cag ccc att gtc acc      1228
332 Gly Glu Glu Val Val Leu Ile Asn Ser Ala Gly Gln Pro Ile Val Thr
333          345          350          355
335 gat gaa tat ggc gtc gct ctt cac cct ctt caa cac gga gcc tcg tac      1276
336 Asp Glu Tyr Gly Val Ala Leu His Pro Leu Gln His Gly Ala Ser Tyr
337 360          365          370          375
339 tat ctg atc tagtcgtgtg ggagatttga gtttgaagaa gaaattaaga      1325
340 Tyr Leu Ile
343 cctgtctctt tctttcacca tctactcgta cgtaggctta aatgttaaga ttttataaag      1385
345 tattggtttc agttacctgt tgtgacggtg tttatgtatg agtttcggac aacattcaca      1445
347 aaactctctc gttaaattgt tgacctaata atatatgatg tgtgtttcat tattaaaaaa      1505
349 aaaaaaaaaa aaa      1518
353 <210> SEQ ID NO: 4
354 <211> LENGTH: 378
355 <212> TYPE: PRT
356 <213> ORGANISM: Arabidopsis thaliana
358 <220> FEATURE:
360 <223> OTHER INFORMATION: Amino acid sequence coding for a protein involved in
361 differentiation
363 <400> SEQUENCE: 4
365 Met Ala Ser Ser Asn Arg His Trp Pro Ser Met Phe Lys Ser Lys Pro
366 1          5          10          15
369 His Pro His Gln Trp Gln His Asp Ile Asn Ser Pro Leu Leu Pro Ser
370          20          25          30
373 Ala Ser His Arg Ser Ser Pro Phe Ser Ser Gly Cys Glu Val Glu Arg
374          35          40          45
377 Ser Pro Glu Pro Lys Pro Arg Trp Asn Pro Lys Pro Glu Gln Ile Arg

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/787,737

DATE: 07/19/2001

TIME: 12:21:33

Input Set : A:\001560-393.ST25.txt

Output Set: N:\CRF3\07192001\I787737.raw

L:11 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:466 M:283 W: Missing Blank Line separator, <220> field identifier
L:479 M:283 W: Missing Blank Line separator, <220> field identifier
L:492 M:283 W: Missing Blank Line separator, <220> field identifier
L:505 M:283 W: Missing Blank Line separator, <220> field identifier